QUANTUM THEORY OF LASER COHERENCE AND NOISE*

Victor Korenman
Lyman Laboratory, Harvard University, Cambridge, Massachusetts
(Received 6 January 1965)

We would like to report some results of a calculation of the properties of a laser which shed some light on the relationships between models which assume a pure sinusoidal output[1] and those which are based on a signal with a finite frequency width.[2]

We employ a fully quantum mechanical description of the laser in terms of correlation functions of the quantized radiation and matter fields. A generalization of the theory of thermal Green's functions[3] allows us to consider a system whose equilibrium is dynamic rather than static. The usual equations for the field propagators {e.g., $\eta(t-t')\langle[E(rt),E(r't')]\rangle$} are supplemented by generalized detailed balancing conditions[4] which yield the self-consistent level occupation densities and the intensities of field fluctuations. For concreteness we consider a gaseous laser in which the cavity is idealized by boundary conditions and a uniformly distributed quantum mechanical loss mechanism. We also treat the pump quantum mechanically.

In a first calculation we assume the existence of a sinusoidally varying mode of the electromagnetic field. We expand the medium polarization in a power series in the field expectation value $\langle E(rt)\rangle$ and, ignoring any large field fluctuations which might be implied, we obtain identically the equations of Lamb, which determine the frequencies and intensities of the stable modes in a self-consistent manner. However the noise in the vicinity of this frequency is infinite. We stress that this is not an accidental property of the model or of the approximation. Considerations similar to those of Goldstone[5] tell us that the frequency tranform of the field correlation function must have a δ-function singularity at any frequency where the system will support a true mode.[6] In the laser such a singularity leads directly to an infinite output power. This is physically inadmissible, and we conclude that $\langle E(rt)\rangle$ is, in fact, zero. The noise, or field fluctuation near the resonant frequency, must build up as the inversion approaches threshold for the stability of a pure mode and must modify the inversion so that such a threshold is not reached.

We next recalculate the laser properties, setting $\langle E(rt)\rangle$ equal to zero but allowing for the presence of large narrow-band field fluctuations, perhaps at several frequencies. For an inversion larger than Lamb's threshold we find that new self-consistent solutions exist with values of the intensities and frequencies which differ negligibly from those obtained by "mode theory" (the calculation with $\langle E\rangle$ not zero). In the present approach, however, the linewidth is determined. In the particular case of a single traveling wave mode, and to the order of approximation of Lamb's calculation, the width may be expressed as

$$\Delta f=\frac{2\pi h\nu_0(\Delta\nu^C)^2}{\text{power output}}\left[\frac{N_2}{N_2-N_1}+\frac{N_2+N_1}{N_2-N_1}d\right],$$

where Δf is the full width at half-maximum, ν_0 is the center frequency, N_2 and N_1 are, respectively, the unsaturated values of the upper- and lower-state occupation densities, $\Delta\nu^C$ is the cavity half-width, and d is related to the field intensity. In Lamb's notation, $d=I_n\mathfrak{N}/4$.[7] This expression is half the Townes[2] width with the correction $N_2/(N_2-N_1)$ previously calculated by Shimoda.[8] The last term is a saturation effect and is small since d must be small. Gordon[9] has recently found a similar expression.

The reason for the agreement of the above calculation with that of Lamb is easy to see. The nonlinear effects which lead to saturation and mode interactions come from the effect on the active atoms of the laser signal which is described by the field fluctuation $\langle E(rt)\times E(r't')\rangle$. An intense narrow-band output is described by a large narrow peak in the frequency transform of this function (or several such peaks for multimode operation). Lamb, in effect, has replaced this fluctuation by $E^2\cos\omega_0(t-t')$ (or, equivalently, its frequency transform by δ-function peaks at $\pm\omega_0$). If the width of the peak we find is small compared to the natural linewidths of the atomic levels, calculations will not distinguish between such a peak and a δ function. This is merely the mathematical reflection of the fact that the time of the atomic processes is short compared to the resolving time for the almost monochromatic electric field. Thus, in the case of a narrow peak, our second expression for the correlation function will have the same dependence on mode frequencies and intensities as the first or "mode theory." Instead of finding self-consistent modes, however, we determine values of the operating characteristics for which self-consistently calculated correlation functions reproduce their assumed narrow peaks. The self-consistent values of the intensity and "complex frequency" when the imaginary part of the frequency, the width, is small, can differ only slightly from the mode-theory intensity and real frequency.

Although our calculation in which true modes do not exist produces the results Lamb obtains in "mode theory" and, in addition, gives a consistent description of the fluctuations, it is desirable to perform the mode calculation as a first approximation. The formal reason is that the Goldstone theorem which insures the correct position of the peaks of the correlation function is not true in an expansion to any given order of perturbation theory. Higher order terms, important only near the resonances, must be included in any approximation to satisfy it, and agree with the equivalent order calculation in mode theory. The mode-theory calculation provides a guide for selecting these terms.

As a final remark we note that the absence of a pure mode does not mean the absence of interference effects between two separate lasers. The intensity correlation function[10] appropriate to the field produced by two independent sources is identically given by

$$\begin{aligned}&\langle I(x')I(x)\rangle-\langle I(x')\rangle\langle I(x)\rangle\\ &\equiv\langle[E_1^-(x)+E_2^-(x)][E_1^-(x')+E_2^-(x')][E_1^+(x')+E_2^+(x')][E_1^+(x)+E_2^+(x)]\rangle\\ &\quad-\langle[E_1^-(x)+E_2^-(x)][E_1^+(x)+E_2^+(x)]\rangle\langle[E_1^-(x')+E_2^-(x')][E_1^+(x')+E_2^+(x')]\rangle\\ &=\langle E_1^-(x)E_1^-(x')E_1^+(x')E_1^+(x)\rangle-\langle E_1^-(x)E_1^+(x)\rangle\langle E_1^-(x')E_1^+(x')\rangle+\langle E_2^-(x)E_2^-(x')E_2^+(x')E_2^+(x)\rangle\\ &\quad-\langle E_2^-(x)E_2^+(x)\rangle\langle E_2^-(x')E_2^+(x')\rangle+\langle E_1^-(x)E_1^+(x')\rangle\langle E_2^-(x')E_2^+(x)\rangle\\ &\qquad+\langle E_2^-(x)E_2^+(x')\rangle\langle E_1^-(x')E_1^+(x)\rangle,\end{aligned}$$

where E_1 is the field produced by the first laser, and E_2 that produced by the second. The superscripts refer to positive- and negative-frequency components of the fields, and (x) is an abbreviation of (rt). If the intensity fluctuations of the individual laser outputs (the first two pairs of terms after the equality sign above) are small,[11] the remaining term dominates. If, further, we can consider the output of each laser as dominantly a single spatial mode $g_i(r)$ with a narrow frequency width $2\delta_i$ centered at ω_i, the field correlation function $\langle E_i^-(x)E_i^+(x')\rangle = I_i g_i^*(r)g_i(r')\exp[-i\omega_i(t-t')-\delta_i(t-t')]$. The intensity correlation function is

$$I_1I_2g_1^*(r)g_1(r')g_2^*(r')g_2(r)$$

$$\times\exp[-i(\omega_1-\omega_2)(t-t')-(\delta_1+\delta_2)(t-t')]\text{ plus c.c.}$$

This is precisely the intensity correlation one would find by performing a time average in the case of classical waves with spatial distributions $g_1(r)$ and $g_2(r)$ and frequencies ω_1 and ω_2, multiplied by the exponential decay term. In particular, for plane-wave modes it is

$$\langle I(x)I(x')\rangle-\langle I(x)\rangle\langle I(x')\rangle$$

$$=I_1I_2\cos[(k_1-k_2)(r-r')-(\omega_1-\omega_2)(t-t')]$$

$$\times\exp[-(\delta_1+\delta_2)(t-t')].$$

We can interpret this as a fully visible diffraction pattern at any time, with a slow random drift of the position of any particular feature characterized by a relaxation time $(\delta_1+\delta_2)^{-1}$. Thus, contrary to previous assertions[12] such a fully visible pattern is obtainable from the interference of two lasers even though $\langle E\rangle = 0$, as long as the intensity fluctuations of the individual lasers are small, and is practically observable when the linewidths are sufficiently small.

We would like to thank Peter Schweitzer and especially Professor Paul Martin for many enlightening discussions.

*Work supported in part by the National Science Foundation.

[1]W. E. Lamb, Jr., Phys. Rev. 134, A1429 (1964).

[2]A. L. Schawlow and C. H. Townes, Phys. Rev. 112,

1940 (1958); and many other authors.

[3]P. C. Martin and J. Schwinger, Phys. Rev. 115, 1342 (1959).

[4]J. Schwinger [J. Math. Phys. 2, 407 (1961)] has obtained similar conditions. Our work is a generalization of his.

[5]J. Goldstone, A. Salam, and S. Weinberg, Phys. Rev. 127, 965 (1962).

[6]Application of the same considerations to the theory of the superfluid yields the Pines-Hugenholtz theorem, which states that the superfluid excitation spectrum cannot have a gap. See P. C. Hohenberg, thesis, Harvard University, 1962 (unpublished); N. M. Hugenholtz and D. Pines, Phys. Rev. 116, 489 (1959).

[7]See reference 1. $\mathfrak{N}$ is defined by Eq. (84) and I_n by (98).

[8]K. Shimoda, Inst. Phys. and Chem. Res. (Tokyo) Sci. Papers 55, 1 (1961); K. Shimoda, H. Takahasi, and C. H. Townes, J. Phys. Soc. Japan 12, 686 (1957).

[9]E. I. Gordon, Bell System Tech. J. 43, 507 (1964), especially Eq. (46). Note that his level densities are saturated values and that he has a homogeneously broadened line while ours is inhomogeneously broadened as well.

[10]For a discussion of the importance of such quantities see R. J. Glauber, Phys. Rev. 130, 2529 (1963). L. Mandel [Phys. Rev. 134, A10 (1964)] has a discussion similar to the following but does not stress the consequences of small intensity fluctuations of the individual sources.

[11]As Gordon (reference 9) has stressed, these fluctuations must be small due to the effects of saturation. That their space average is small for large time differences is evidenced by the almost sinusoidal beat note that has often been observed on mixing two laser outputs. That their time average is small for macroscopic space separations is evidenced by the fact that the focusing properties of a single laser output are consistent with the assumption that the output is a classical field in a single mode. For a thermal source the intensity fluctuations are not small. They are, for example, entirely responsible for the Brown-Twiss effect. A direct proof of the smallness of these terms in the laser requires an analysis of intensity correlation function. A discussion of this will be published elsewhere, along with a full account of the calculations described above. H. Haken [Phys. Rev. Letters 13, 329a (1964)] has claimed to prove that the intensity fluctuations are indeed small. However, there are some mathematical problems in his decomposition of the field operators [see L. Susskind and J. Glogower, Physics 1, 49 (1964)].

[12]T. F. Jordan and F. Ghielmetti, Phys. Rev. Letters 12, 607 (1964). We do not really disagree. One might say that the field expectation value is nonzero "at any particular time," but that random fluctuations make it impossible to predict its phase with any confidence at a time more than an inverse linewidth later.